White Paper Report

Report ID: 98505

Application Number: HD5065009

Project Director: Amy Gansell (amygansell@yahoo.com)

Institution: Unaffiliated Independent Scholar

Reporting Period: 4/1/2009-10/31/2010

Report Due: 1/31/2012

Date Submitted: 1/31/2012

WHITE PAPER

HD5065009

Identifying Regional Design Templates of Ancient Near Eastern Ivory Sculptures of
Women Using Computer Technology

Amy Rebecca Gansell, Project Director
Independent, Unaffiliated

1/31/12

Background and Accomplishments

While not reaching the goals specifically articulated in the proposal, research has revealed that those goals were unattainable as strictly defined. However, the project has produced much more nuanced results, which are in fact more meaningful in relation to the corpus of ancient artworks investigated.

For nearly a century, scholars have puzzled over the classification and regional attribution of ancient Near Eastern ivory sculptures – the problem lies in determining "which statues came from where." Dated to the early first millennium BCE, almost all of these small sculptures, which served as decorative components of luxury objects and furniture, were found in northern Iraq, where they had been imported in antiquity from their Levantine contexts of production in Syria and Lebanon. Because there is very limited textual and archaeological evidence for classifying them by region and/or decade of production, art historical attempts at classification have relied on the visual observation and subjective evaluation of their style and physical characteristics.

This project, "Identifying Regional Design Templates of Ancient Near Eastern Ivory Sculptures of Women Using Computer Technology," was initiated to offer a comparative quantitative approach to the art historical classification of Levantine ivory sculptures. Focusing on a corpus of carved female figures, it employs computer science techniques and assesses the sculptures through both previously considered, visually recognizable data (such as type of hairstyle) and newly recorded proportional data reflecting physiognomic proportions (such as the ratio of the height of an eye to the length of the face), which were calculated based on measurements I took during first-hand study of the sculptures between 2004 and 2006. In total, we had approximately 30,000 pieces of data to work with – far more than could be effectively sorted through manual or traditional methods.

The overall objective of this project was to identify quantitatively derived templates reflecting the conceptual prototypes motivating artists of various regions in the production of ancient Near Eastern carved ivory figures of women. Through computer-aided sorting and grouping, the results were intended to aid in the regional classification of the sculptures as well as to reveal ancient aesthetics and ideals of feminine beauty. Contrary to the expectations stated in my proposal, it is not possible to identify templates or formulate typological prototypes. However, we have clarified the factors that influence each artistic outcome and found relationships among them that are relevant to both classification and aesthetics. Proportional, technical, iconographic, and qualitative variables contribute to the appearance of the sculptures, whose variously configured features can occur within *or outside* of standard, recognizable ranges of variation. The skill of the carver, the care with which the object was carved, and the physical properties of the ivory itself also affect artistic outcome.

In terms of classification and aesthetics, our analysis has generally correlated to and confirmed art historical interpretations. But we have also elucidated information that would be impossible to retrieve through human observation alone.

First, we have quantitatively demonstrated the most typical proportional ranges for facial features on these sculptures. In so doing, we have clarified that proportional relationships on the vertical axis of the face are generally more consistent and were probably more important than horizontal proportions in carving a well executed, culturally attractive figure. In addition, we can identify which variables (proportional or iconographic) are most statistically significant in the grouping of the sculptures. These variables, as well, might have been more significant to the ancient carvers and consumers of these works.

Second, we have been able to identify, within classification groups, how likely the membership of each object is to that group; this tells us how representative of that group the object is. For example, an object with 90% likelihood of membership to a group is more representative of that group than one that has only a 30% likelihood of membership (which itself would have a higher likelihood of membership to another group). In cases when multiple objects demonstrate the same primary and secondary membership likelihoods, we might see a relationship or overlap between the groups. We might also have revealed a sub-group of hybrid sculptures (a point of interest to art historians).

Furthermore, our research produced an unexpected art historical and archaeological outcome. Observing their tendencies to group and the relationships among them, it appears that sculptures excavated from the sites of Arslan Tash and Khorsabad are more closely related than those excavated from the site of Nimrud. This suggests that, differing from circumstances at Nimrud, the sites of Arslan Tash and Khorsabad might have received sculptures from the same or closely related sources or during the same or closely related historical phases. This information encourages further historical research that might clarify the ancient relationship among these sites.

While more research is necessary, some of our algorithms appear to have the potential to predict how a fragmentary object might be reconstructed. There is promise, too, that we have identified some sub-groups of objects that might represent separate production centers within a region or separate historical production phases. Variations in iconographic details such as the manner in which a headband is depicted, however, seem to be idiosyncratic. That is, there does not appear to be any pattern to the variation; rather, the differences can be attributed to the will or whim of the artist.

Project Activities (4/1/2009-10/31/2011)

Project activities can be divided into three phases: 1) an initial phase of research and planning, 2) a first attempt at implementation, and 3) following a one-year extension, a second approach to implementation employing a different skill-pool of student researchers and programmers.

PHASE 1 (4/2009-7/2009)

The initial phase of this project, from March to July 2009, entailed consultations with my

co-Principle Investigator, Chris Wiggins, Associate Professor of Applied Mathematics at Columbia University. We further researched and assessed the best-suited computing techniques, which, in turn, clarified the skill, experience, and time-commitment we would look for in student programmers. We decided to approach our project as a grouping problem, meaning that we would ask the computer to separate our assemblage of archaeological objects according to relationships it identified among the unique matrices of data for each object. These data matrices included proportional data (such as the ratio of the height of an eye to the length of the face) and visual data (such as type of jewelry and hairstyle). Once results were achieved, I, an art historian, in consultation with the art historian on my advisory board, would interpret them in terms of aesthetics and the regional classification of ancient Near Eastern art.

During this first project phase, we also gave careful consideration to "what will art historians and archaeologists want to know." This resulted in a "mission statement" for the project that we could return to in order to assess the humanistic utility of the research underway. Indeed, this kept what became a complex and often abstract computer science exploration (in the subfields of data mining and machine learning) focused on attaining results that would contribute to humanistic scholarship.

During this planning phase, I attended two conferences in order to see what other projects were currently working at the crossroads of art history, archaeology, and computer science. In addition to keeping abreast of related projects, I was able to learn what types of software languages and environments were most commonly used and what pitfalls were most commonly encountered (not enough data and unedited data). The first conference I went to was the annual international meeting of the Computer Applications and Quantitative Analysis in Archaeology (CAA) in Williamsburg, VA, March 22-26, 2009. I received permission to attend this as part of the funded project, although it took place a week before the standard April 1 project start-date. The second conference I attended was the annual international Digital Humanities meeting held at the University of Maryland, June 22-25, 2009.

PHASE 2 (7/2009-10/2010)

In mid-July 2009, the first implementation phase of this project began. We immediately advertised for Master's level students to help with programming, but it was not until the academic semester started, that in October 2009, we were able to retain the commitment of two student programmers. It proved more difficult than expected to find and retain student support, because the skill-set was highly specific and the necessary time-commitment, on top of their required coursework, was challenging for most full-time students, unless they could receive independent research credit for project participation.

In order to focus time and funding on the project's primary methodology (that is, a computer science approach to archaeological and art historical classification), I altered my proposed research plan. Originally I had planned to compare the results of a computer science approach with results produced by more traditional statistical methods; this, however, would have entailed the unmanageable scenario of simultaneously

4

assembling and directing two groups of student programmers under two different faculty members. Although the statistical approach would have provided a useful and interesting comparison, it was not necessary to validate the accuracy of the computer science results. Therefore, I proceeded only with a computer science team.

As the first phase of project implementation got under way, we found that our biggest challenge was the nature of the data itself. Dealing with fragmentary archaeological objects, my dataset was literally full of holes – it is riddled with missing data that cannot be physically recuperated. Although preliminary tests prior to the grant period suggested that we could proceed with an established algorithm, on account of the missing data, this turned out not to be the case. So we went back to the drawing board, researching and experimenting with algorithms best suited to the fragmentary corpus. We arrived at a solution in a reliable and fully functional algorithm, but it only treated what is known as the "real" data (that is the numerical, proportional data). The "categorical" data (visually observed descriptive data, such as the type of jewelry depicted on a figure) would need to be handled separately. Once the real and categorical data was fully analyzed and reviewed, we hoped to combine their results in order to make final interpretations.

During this first phase we also developed a web application to visually display the algorithmic grouping of the sculptures in ranked order from "most to least likely to belong to this group." This required the manual scanning and labeling of hundreds of photographs of the objects, but the resulting web app saved an incredible amount of time. I, the art historian, no longer need to arrange my print photographs in groups to 'see' which objects the computer grouped. The web app also made my work more mobile and handy for demonstrations.

By the end of the summer in 2011, I was facing an urgent need for more student staff, as the programming was very time consuming, and we were only about half way to meeting project goals. Because preliminary results were promising, I applied for and was granted an extension of an additional twelve months, from October 31, 2010 to October 31, 2011.

PHASE 3 (10/2010-10/2011)

The third and final phase of this project began during the autumn of 2010. At this juncture, I reassessed the project goals and accomplishments and prepared to establish a fresh team of researchers and programmers to see the project to completion. I also employed a student, who used Django, an open source framework for developing websites, to repair and improve the web application for viewing grouping results.

This time, instead of hiring and tasking individual students, Wiggins established as a weekly research workshop, held at Columbia University. This attracted students who were on an academic track and for whom the possibility of contributing to a publication was valuable. I participated in all meetings, which were overseen by Wiggins, and led by a Postdoctoral Researcher, who, in turn, supervised a group of MA and PhD students. This arrangement resulted in a dynamic and dedicated team, some members of which worked full-time on the project from June into September of 2011.

First, we reconfigured the data by collapsing and lumping variables so that we had a minimum amount of missing data. That is, instead of working around the holes in the data as we had in our first effort of implementation, we sought to minimize the holes from the get-go. During this process, we also discovered and corrected a few cases in which variables, which were dependent on one another (for example, "type of curls" is dependent upon whether a sculpture portrays curly hair), had not been treated as such and had potentially skewed earlier results.

Previous analysis was based only on the real-valued attributes of each object (i.e., numerical values such as the ratio of the height of an eye to the length of the face). This was undesirable, as a lot of highly descriptive information is contained in the categorical attributes (e.g. hair style, or the presence of a necklace) that should not be excluded from analysis, and there is risk for error in trying to re-combine results of separate analyses of real and categorical data. Moreover, from an art historical perspective, real and categorical attributes would not have been separated in the minds and eyes of the people who produced and viewed these objects, so any statistical analysis should ideally integrate both sources of data.

For this reason, the new team started by deriving a method that made it possible to group objects based on both their real and categorical attributes. This method, like the last, relies on a method known as expectation maximization. This determines the group membership of each object, as well as the group averages of each of the attributes. To facilitate interpretation of the results of our model, we ran further analyses to determine which of the many attributes in our dataset most strongly motivated the group membership an object. This was done by calculation of a quantity known as the mutual information. We also ran tests on smaller groups of objects to reveal whether variations within single variables (such as the different designs of headbands) motivated sub-groups that, for example, might be interpreted as characteristic of a specific production workshop.

To gain insight into the accuracy of our method, we employed several types of tests. First, we performed analysis on a pseudo-dataset (or "toy"-dataset), generated by hand, in which each object had a known group membership. Since we knew the "correct" result for this dataset, we could use it to assess the reliability with which group membership could be determined. Second, we constructed a set of randomly generated "synthetic" objects, which had the same range of possible attributes as the objects in our actual dataset. The attribute values for each object were generated in accordance with the averages for each group obtained from analysis of the actual data. Like with the toy dataset, this allowed us to see how the accuracy of our algorithm deteriorates as we strip a larger and larger fraction of object attributes from the data. That is, it could track the increase of accuracy versus the decrease in missing data. Finally, determination of the most appropriate number of groups is a non-trivial task for these types of methods. The previous analysis, based only the real-valued attributes, suggested that the dataset naturally separated into three groups. This result was obtained using a technique known as cross-validation, but our final test determined that cross validation is unreliable in

determining the correct number of groups in datasets where many attributes are missing.

The last stage of our work entailed the development of visual means through which we could demonstrate results. We prepared two-dimensional diagrams suited for static contexts, such as print publication. We also developed a dynamic web page (written in javascript) through which a public user can explore our results using features such as interactive sliders and an indexed image bank.

The only technical revision to the original proposal is that instead of coding in weka, we used python code, which offered the benefit of an extensive set of libraries. Like weka, python is an open source programming language. A key aspect of our open-source approach is that we were able to release all source code for the benefit of other researchers.

Long Term Impact, Audiences, and Continuation of the Project
This project has bridged academic and intellectual relationships between humanities and science. It has exposed a number of computer science students to ancient art and archaeology, demonstrating their potential contributions to the humanities. Likewise, it has opened the eyes of humanities scholars to the information computers can retrieve about human culture that we as humans cannot manually access.

Because results thus far have been so complex and interpretations have been provisional, we have not yet presented the outcome of this project to public or academic audiences. However, I anticipate doing so in the upcoming year. Further art historical analysis of the grouping results is necessary in order to offer final interpretations. I plan to pursue this as part of my own scholarly agenda over the upcoming several months (spring semester 2012). I hope to publish our final results and interpretations in a peer-reviewed journal.

Finally, I have recently begun some long-term research for which I will eventually examine "Near Eastern" ivory sculptures found in Greek contexts, which I believe to be of Mediterranean, not Near Eastern, origin. I plan to test the classification of these "Near Eastern" ivory sculptures of women by comparing their data to that of the corpus of bona fide Near Eastern objects analyzed in this project. I introduced this project and presented my preliminary hypotheses at the Southeast College Art Association meeting in November 2011.

Grant Products
Interactive public web site: http://ane-ivory-clusters.appspot.com/

Splash page
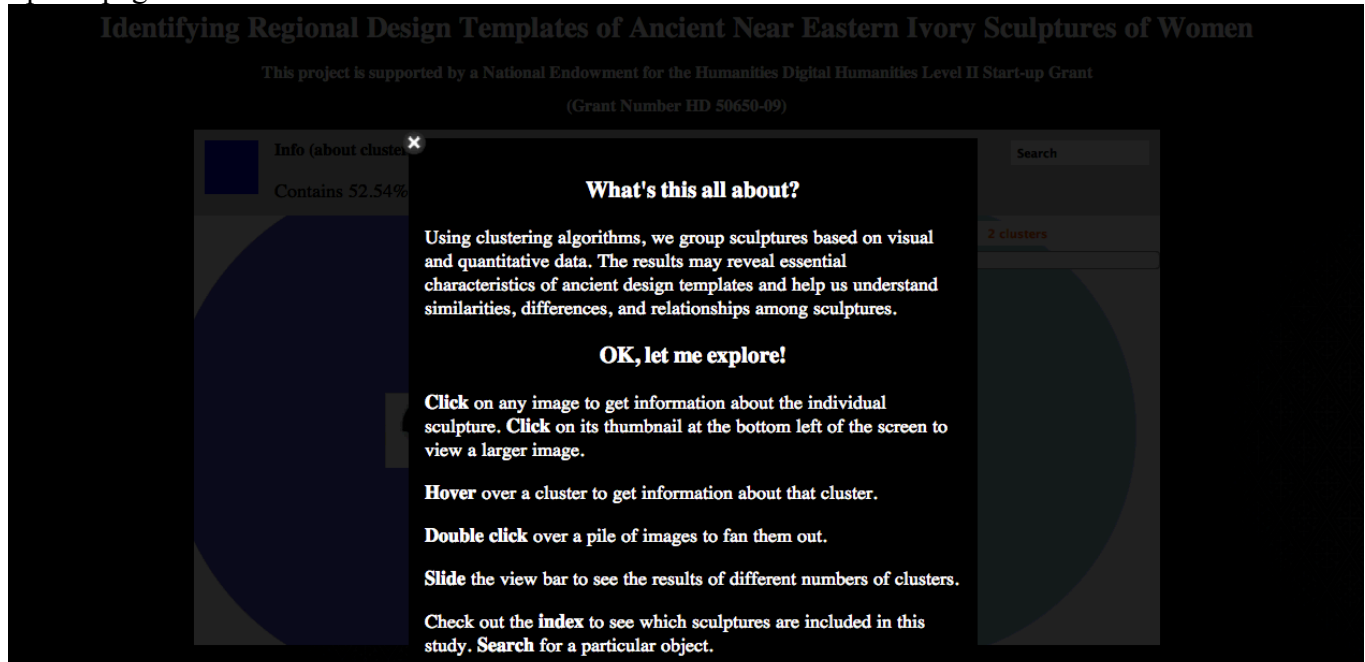
Sample page, showing relationship of objects among groups